GENETICS AND GENOMICS

Narrowing the *wingless-2* mutation to a 227 kb candidate region on chicken chromosome 12

A. E. Webb,[*,2,6] I. A. Youngworth,[*,6] M. Kaya,[†,3] C. L. Gitter,[*,4] E. A. O'Hare,[*,5] B. May,[*] H. H. Cheng,[†] and M. E. Delany[*,1]

[*]*Department of Animal Science, University of California, Davis 95616; and* [†]*USDA-ARS, Avian Disease and Oncology Laboratory, East Lansing, MI 48823*

ABSTRACT *Wingless-2 (wg-2)* is an autosomal recessive mutation in chicken that results in an embryonic lethal condition. Affected individuals exhibit a multisystem syndrome characterized by absent wings, truncated legs, and craniofacial, kidney, and feather malformations. Previously, work focused on phenotype description, establishing the autosomal recessive pattern of Mendelian inheritance and placing the mutation on an inbred genetic background to create the congenic line UCD Wingless-2.331. The research described in this paper employed the complementary tools of breeding, genetics, and genomics to map the chromosomal location of the mutation and successively narrow the size of the region for analysis of the causative element. Specifically, the *wg-2* mutation was initially mapped to a 7 Mb region of chromosome 12 using an Illumina 3 K SNP array. Subsequent SNP genotyping and exon sequencing combined with analysis from improved genome assemblies narrowed the region of interest to a maximum size of 227 kb. Within this region, 3 validated and 3 predicted candidate genes are found, and these are described. The *wg-2* mutation is a valuable resource to contribute to an improved understanding of the developmental pathways involved in chicken and avian limb development as well as serving as a model for human development, as the resulting syndrome shares features with human congenital disorders.

Key words: SNP genotyping, limb development, candidate gene, mutation

INTRODUCTION

The *wingless-2* (**wg-2**) mutation was originally identified at the University of Connecticut in a flock of chickens synthesized from Rhode Island Red, White Leghorn, Dorking breeds, and Red Jungle Fowl (Zwilling, 1956). The *wg-2* mutant phenotype is characterized by the complete absence of forelimbs, severely truncated or absent hindlimbs, craniofacial defects (maxillary region clefting), malformed kidneys, and abnormal plumage (Zwilling, 1956; Somes, 1990; Pisenti et al., 1999). The mutation is inherited in an autosomal recessive pattern and is an embryonic lethal (Somes, 1990; Pisenti, 1995). Reciprocal tissue transplantations between mesoderm and ectoderm suggest that the *wg-2* mutation affects ectoderm (Carrington and Fallon, 1984; Hamburger and Hamilton, 1992). *Wg-2* was the second wingless mutation characterized in the chicken; the first, now extinct, exhibited a similar phenotype and additionally was described as having incomplete or absent lungs, air sacs, and kidneys (Waters and Bywaters, 1943). An improved understanding of the *wg-2* developmental mutation will lead to new knowledge of important developmental pathways in poultry affecting limb, craniofacial (e.g., beak), kidney, and feather development. Further, the *wg-2* developmental mutation is an important model system for the investigation of genes and pathways involved in human syndromes with craniofacial (cleft palate), limb (tetra-amelia, phocomelia) and organ (kidney) malformations.

In the late 1980s at the University of California Davis (**UCD**), a congenic inbred line was established by introgression of *wg-2* in order to advance mapping and gene analysis research at a time wherein genomic tools were barely envisioned, much less the advent of the chicken sequencing project to advance poultry and vertebrate biology. The Wingless-2.331 (**Wg-2.331**) congenic inbred line was derived by first crossing known *wg-2* carriers (*+/wg-2*) into the UCD 331.003 congenic line (Abplanalp et al., 1992; Pisenti et al., 1999). The *+/wg-2* carrier progeny were confirmed by test-mating and then used to backcross to the parent line UCD 331.003. The backcross method using confirmed carrier (*+/wg-2*) progeny was repeated for at least 7 generations, with 2 additional backcrosses after the line was closed to improve fertility and viability, as per available breeding records and personal communications. This labor-intensive multi-year strategy placed the *wg-2* locus in a uniform genetic background (99+% inbred) such that only the introgressed region containing the mutation along with other normal polymorphisms differs between normal (*+/+* or *+/wg-2*) and mutant (*wg-2/wg-2*) individuals. The line was closed (inter se matings only) and perpetuated using *+/wg-2* parents as breeders after being verified by test crosses. Over more than 60 yr of breeding, Mendelian ratios fit an autosomal recessive pattern of inheritance for the *wg-2* mutation.

Here, we describe genetic and genomic analyses of the congenic inbred line Wg-2.331 that have significantly advanced our knowledge of the genomic region incorporating the *wg-2* mutation—work that took place over a decade and spanned 4 chicken genome builds. The work is presented in galGal5 coordinates for the sake of clarity, with references to the state of the build at the time when necessary. At the outset, the chromosomal location and subregion of the *wg-2* mutation was mapped by using a custom Illumina 3 K SNP chicken array (Muir et al., 2008a) wherein a tightly linked SNP was identified, which had the advantage of being predictive of carrier status. Subsequently, new SNP were identified and used for fine-mapping, which, combined with exon-based sequencing of candidate genes within the region, significantly narrowed the causative region.

MATERIALS AND METHODS

Genetic Lines

All birds are derived from the UCD Wg-2.331 developmental mutant congenic inbred chicken line. The birds initially carrying the *wg-2* mutation were bred into the UCD 331.003 line, which is congenic with UCD 003 [a highly inbred line (F > .99)] except at the major histocompatibility complex (Abplanalp et al., 1992) and a small number of other SNP loci (Robb et al., 2011). Samples throughout this paper are designated as wild type *+/+* (normal phenotype), carrier *+/wg-2* (normal phenotype), or mutant *wg-2/wg-2* (abnormal phenotype). The birds used in this research were cared for by trained staff and according to a protocol approved

Table 1. TaqMan® primer-probe sequences for the SNP390 (rs14034687) genotyping assay used to predict UCD Wg-2.331 genotypes: *+/+, +/wg-2, wg-2/wg-2*.

SNP coordinate[1]	Chr 12: 5,061,391
Amplicon coordinates[1]	Chr 12: 5,061,327 - 5,061,449
Forward primer	5′ CTCTTTTGTTAGAGCTGTGCTATG-G 3′
Reverse primer	5′ CAGGACTGTTGCTTTTGTTCTT-TATAGTTTT 3′
VIC probe (*+/+*)	5′ CATGAACA*C*AAAGCC 3′
FAM probe (*wg-2/wg-2*)	5′ CATGAACA*T*AAAGCC 3′

[1]Coordinates based on Dec 2015 Gallus_gallus-5.0/galGal5 chicken assembly for GGA 12.

by the UC Davis Institutional Animal Care and Use Committee (Protocol #18816).

Illumina 3,072 (3 K) SNP Array Analysis

Samples were collected from 9 adult Wg-2.331 birds confirmed as carriers (*+/wg-2*) by test-mating and 11 phenotype-confirmed mutant embryos (*wg-2/wg-2*) at 10 d of embryogenesis (**10E**). DNA was isolated and purified from blood stored in heparin (adults) or sodium citrate (embryos). All DNA samples were isolated using QIAamp® Blood Mini Kit (Qiagen, Valencia, CA) and sent to Illumina, Inc. (San Diego, CA) for analysis. Samples were tested for 3,072 SNP evenly spaced throughout the chicken genome (Gitter, 2006; Muir et al., 2008a; Muir et al., 2008b). SNP linked to *wg-2* should be homozygous in all *wg-2* mutants, heterozygous in *wg-2* carriers, and homozygous for the alternative allele in UCD 331.003.

Genotyping Wg-2.331 Birds Using a Linked SNP

The 3 K SNP analysis found one SNP (rs14034687 aka SNP-16–390-35,608-S-1 and referred to as "SNP390") in complete linkage disequilibrium (**LD**) with the mutant phenotype and, thus, had high value in screening for *wg-2* status. An Applied Biosystems Custom TaqMan® SNP Genotyping Assay was developed for this purpose (Table 1). Real-time PCR was performed on DNA samples (10 ng/μl) isolated from Wg-2.331 individuals to determine genotypes for SNP390 using TaqMan® Universal PCR Master Mix (Applied Biosystems, Carlsbad, CA) under standard conditions. After 15 min at 95°C to activate the polymerase, samples underwent 45 cycles of 95°C for 15 s, 60°C for 1 min, and a plate read. Genotypes were confirmed by comparison to previously genotyped (SNP390) or phenotype-confirmed samples representing the 3 forms of the SNP (normal *+/+, +/wg-2*; mutant *wg-2/wg-2*) with test-mating to affirm carrier status (*+/wg-2*). Samples were collected for DNA preparation and genotype determination at one of 3 time points, depending on the experiment: adults (blood), one d post-hatch chicks (pin feather), or embryos (blood).
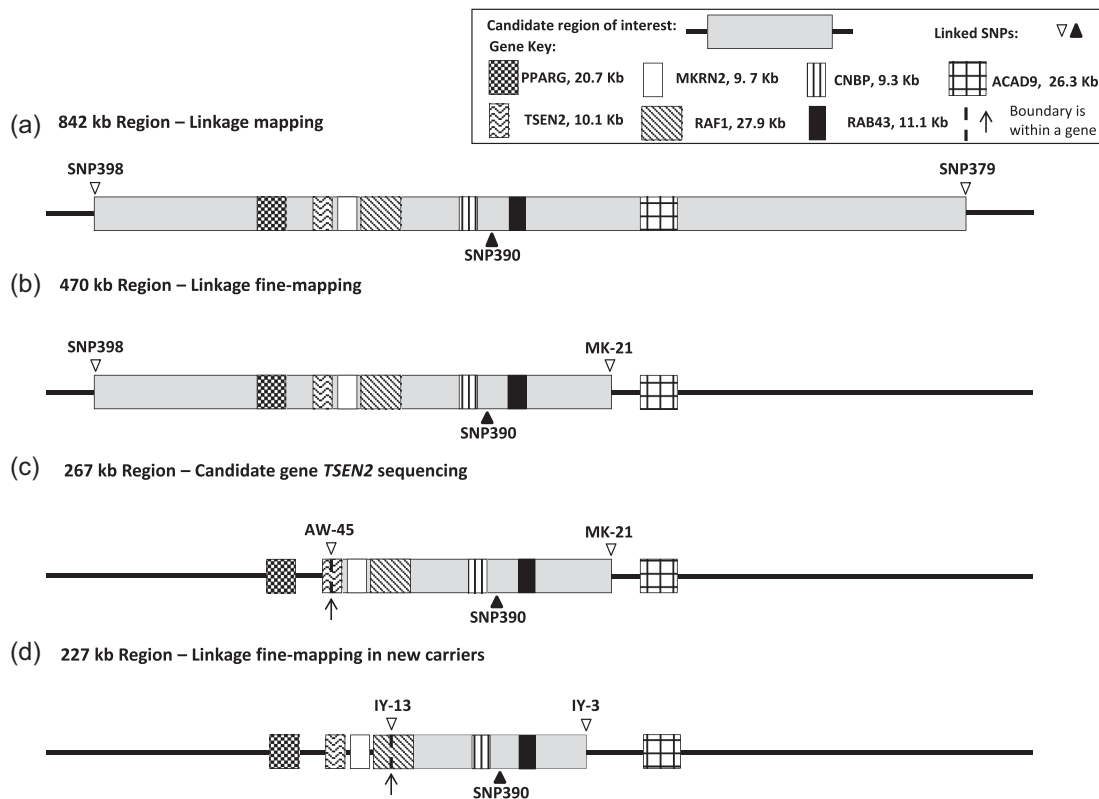
Figure 1. Refinement of GGA 12 *wingless-2* causative region by SNP linkage analysis over multiple generations and genome assemblies. Each gene is denoted by a different shaded box (see Figure 1 Key). The causative region was refined from 842 kb containing 7 annotated genes initially to 227 kb with 3 annotated genes, one partial. The long gray box indicates the causative region. Open triangle heads indicate SNP used for genotyping the margins of a linked region, filled triangle heads indicate the SNP390 genotyping SNP. A dotted line with an arrow indicates that the region boundary exists within a gene (see Supplemental Tables S1–S2, S4–S6 for further information on SNP).

Fine-Mapping

SNP analysis was performed on 79 new individuals to assess the size of the Wg-2.331 causative region segregating in the population. Primer pairs (Supplemental Table S1) were designed (Amplify 1.2, Engels 1993) to amplify 800 to 1,000 bp chromosomal fragments across the region. Eight informative polymorphisms were identified (Supplemental Table S2). DNA was isolated from normal birds ($+/+$, $+/wg$-2) and mutant embryos (wg-$2/wg$-2), and genotypes were determined using SNP390. PCR reactions used GoTaq® (Promega, Madison, WI) under standard conditions. Amplicons were purified by QIAquick® Spin Kit (Qiagen, Valencia, CA) and sequenced by Sanger sequencing on an ABI 3100 (Applied Biosystems, Carlsbad, CA).

Exon Sequencing of Candidate Genes

Exon sequencing was performed on the 6 candidate genes found within the original causative region (WUGSC 1.0/galGal2): *PPARG, TSEN2, MKRN2, RAF1, CNBP,* and *ACAD9* to identify polymorphisms within the exons of each gene. (Figure 1A, Supplemental Table S3). Sequencing was performed using DNA from 6 samples after genotyping with SNP390. Five were from the UCD Wg-2.331 line including one normal homozygote ($+/+$), 2 normal heterozygotes ($+/wg$-

2), and 2 mutant homozygotes (wg-$2/wg$-2). The sixth sample was from UCD 001 (Red Jungle Fowl), the line used for the chicken genome reference sequence (International Chicken Genome Sequencing Consortium, 2004). Sanger sequencing was conducted on an ABI 3100 (Applied Biosystems, Carlsbad, CA).

Candidate Gene TSEN2: Haplotype Analysis

Sanger sequencing was conducted on DNA samples from 4 mutant embryos (phenotype-confirmed and SNP390 genotyped as wg-$2/wg$-2) to investigate their *TSEN2* haplotype. Primer pairs were created using Primer3 (Untergasser et al. 2012; see Supplemental Table S4) to amplify the 10 kb *TSEN2* gene with 2 kb of flanking 5′ and 3′ regions. Amplicons were designed to range in length from 400 to 600 bp and overlapped to resolve poor nucleotide calls. All PCR reactions were performed using Phire™ (Finnzymes, Lafayette, CO) under standard conditions for all primers except AW467–1, -2, and -3, which required 10% DMSO. The *TSEN2* SNP status across one 500 bp region also was investigated in an additional 29 genotype-confirmed mutant embryos. Purified amplicons were sequenced (Davis Sequencing, Davis, CA) and aligned by Sequencher (Gene Codes, Ann Arbor, MI). One embryo

exhibited evidence of a recombination event and was thus also sequenced across the entire *TSEN2* gene.

Linkage Analysis in New Generations of Carriers

After several years of maintaining the line without fine-mapping, the 2014 and 2015 carriers used to perpetuate the line (generating 2015 and 2016 offspring, respectively) were genotyped at up to 19 additional SNP (Supplemental Table S5) to identify recent recombination events. The SNP selected for this purpose were identified from sequence data developed from a capture array experiment employing 2 pooled *wg-2* embryos (Robb and Delany, 2012). These SNP were previously observed in other breeds, shown by their listing in the Beijing Genome Institute database (http://chicken.genomics.org.cn/). Using these SNP precluded problems with validation and variant calling confidence, as they were identified by 2 independent sources. The birds were genotyped at SNP390, test-mated, and then examined at the additional SNP loci to find individuals in which linkage between a variant and *wg-2* carrier status was lost (lack of association), indicating recombination. Initial SNP genotyping assessed linkage at the ends of the region with only a couple of SNP. In the case in which genotypes did not retain their expected heterozygous condition, additional SNP oriented further inward were assessed to determine the breakpoint of the putative recombination event in these birds. PCR reactions used Phire Hot Start II DNA Polymerase (ThermoFisher Scientific, Waltham, MA), and PCR products were sequenced by Sanger sequencing (Davis Sequencing, Davis, CA, or Quintara Biosciences, Berkeley, CA).

RESULTS

Mapping of the Wingless-2 Mutation

To map the mutation to a chromosome and causative region, SNP were screened for linkage with the mutant phenotype using the Illumina 3 K SNP chicken array based on the WUGSC1.0/galGal2 assembly. Linkage analysis was performed in 20 Wg-2.331 individuals consisting of 9 $+/wg$-2 carrier adults and 11 wg-2/wg-2 mutant embryos. Of the 3,072 SNP assayed, 2,733 SNP amplified successfully (Muir et al., 2008a). Twelve SNP associated with GGA 12 and linked to *wg-2* were identified establishing a 7.16 Mb region of interest between 4,180,275 and 11,344,252 bp (Gitter, 2006; Supplemental Table S6). This region in galGal5 is from 4,121,073 to 11,352,192, establishing a 7.23 Mb region of interest. Of these SNP, SNP390 was in complete LD and was used for genotyping all subsequent samples (Table 1). The GGA 12 causative region was estimated to be a maximum size ($\mathbf{CR_{max}}$) of 842 kb after observing recombination involving SNP398 and SNP379 in 3 proven carrier birds ($+/wg$-2) (Figure 1A). This region based on the galGal5 assembly encodes 7 known genes: *PPARG, TSEN2, MKRN2, RAF1, CNBP, RAB43,* and *ACAD9.*

Fine-Mapping

An additional set of 79 Wg-2.331 birds was sequenced across polymorphic regions using new primers (Supplemental Table S1) to enhance marker content across the region and improve linkage analysis. Recombination was observed in 13 individuals. Nine individuals exhibited recombination involving 3 to 4 markers confirming the 5′ CR_{max} boundary of the causative region to marker SNP398, located at 4,724,679 bp on galGal5 (Supplemental Table S5). Five individuals (one of these experienced recombination at both ends of the region) indicated a recombination involving 4 to 5 markers, narrowing the region at the 3′ boundary of the CR_{max} to marker MK-21 located at 5,194,358 bp, inward from marker SNP379. In summary, the *wg-2* causative region was reduced to a CR_{max} of 470 kb spanning 4,724,679 to 5,194,358 bp on galGal5, which eliminated one gene (*ACAD9*) from an original list of 7 candidate genes (Figure 1B).

Exon Sequencing of Candidate Genes

As an alternative approach to identifying the *wg-2* mutation, exons of the 5 remaining validated genes at the time were sequenced (Supplemental Table S3) to uncover potentially causative amino acid substitutions. This work was performed while galGal3 was the operative genome build, and 5 genes were included in the study: *TSEN2, RAF1, MKRN2, CNBP,* and *PPARG.* Sequencing of 5 Wg-2.331 individuals of confirmed phenotype and SNP390 genotype (one normal $+/+$, 2 carriers $+/wg$-2, and 2 mutants wg-2/wg-2) and one Red Jungle Fowl (UCD 001) revealed nucleotide polymorphisms within exons of 3 candidate genes [*TSEN2* ($n = 25$), *RAF1* ($n = 1$), *MKRN2* ($n = 4$)], whereas no polymorphisms were detected in 2 candidates (*CNBP* and *PPARG*). Amino acid predictive analysis revealed that of 25 *TSEN2* polymorphisms, 7 encoded non-synonymous amino acid substitutions, whereas predictions of the polymorphisms in the other 2 genes indicated synonymous substitutions. In silico alignment (see Figure 2) revealed that 3 of the amino acid substitutions (Y149D, A211T, and A214T) were previously reported in the DT40 cell line (NP_001025765.1 (Baba and Humphries, 1984; Caldwell et al., 2005)), suggesting these would not be likely candidates (Figure 2). One substitution differed among mutant, normal (wild type), and DT40 (M356V) genotypes, and 3 other substitutions were unique to *wg-2* mutants (I75V, A232G, and N459S). These results prompted further investigation of *TSEN2* specifically.

```
                                                              (b)
DT40_SEN2           MAEAAFHPPRRKRRVFECYQAPFPVGPGAGDFRMCRADMVNNCVVVRSPEDMDQLYNKGY 60
WG-2.331_MUTANT_SEN2 -----------------------------------------------------------GY 2
WG-2.331_WT_SEN2     -----------------------------------------------------------GY 2
                                                                          **

                                   (b)
DT40_SEN2           FGKGILSRSRPVYSISDPSLVSKWQGANLNVPIITSKKYQQRVQWAKSALQEQGFDGCSV 120
WG-2.331_MUTANT_SEN2 FGKGILSRSRPVYSVSDPSLVSKWQ------------YQQRVQWAKSALQEQGFDGCSV 49
WG-2.331_WT_SEN2     FGKGILSRSRPVYSISDPSLVSKWQ------------YQQRVQWAKSALQEQGFDGCSV 49
                     ************* :*********             ********************

                                                (a) (c)
DT40_SEN2           TKILENYTRPLEFPFLEKGGAEQTGDSQDIVCPNTENTELSGQSSTDTGNIATSSPECQK 180
WG-2.331_MUTANT_SEN2 TKILENYTRPLEFPFLEKGGAEQTGDSQDIMCPNTENTELSGQSSTDTGNIATSSPECQK 109
WG-2.331_WT_SEN2     TKILENYTRPLEFPFLEKGGAEQTGDSQYIMCPNTENTELSGQSSTDTGNIATSSPECQK 109
                     ****************************  *:.:***************************

                                               (a) (a)                    (b)
DT40_SEN2           EVCQKAYAEGDPASDSMVGSKEQEPADVKETAQTSCQMHGSCMHCGCKAKGAPDRESCNM 240
WG-2.331_MUTANT_SEN2 EVCQKAYAEGDPASDSMVGSKEQEPADVKETAQTSCQMHGSCMHCGCKAKGGPDRESCNM 169
WG-2.331_WT_SEN2     EVCQKACAEGDPASDSMVGSKEQEPADVKEAAQASCQMHGSCMHCGCKAKGAPDRESCNM 169
                     ****** *********************** :**.:******************* *******

DT40_SEN2           AKSRERAPEYVLVQEEEESSLCPEEGSAREQENFVKKEKLVCRRNPFRIFEYLQLSLEEA 300
WG-2.331_MUTANT_SEN2 AKSRECAPEYVLVQEEEESSLCPEEGSAREQEN-----------------------A 203
WG-2.331_WT_SEN2     AKSRERAPEYVLVQEEEESSLCPEEGSAREQEN-----------------------A 203
                     ***** ****************************                        *

                          (d)
DT40_SEN2           FFLVYAIGCLTVYYGEEPLTILKLWEIFSEVKPSFKTTYMAYHYFRSKGWVPKVGLKYGT 360
WG-2.331_MUTANT_SEN2 FFLVYAVGCLTVYYGE------------------------------------------- 218
WG-2.331_WT_SEN2     FFLVYAMGCLTVYYGE------------------------------------------- 218
                     ****** :*********

                                                                    (b)
DT40_SEN2           DLLLYRKGPPFYHASYSVIAELVDDNFEGSLRRPLSWMSLSGLNRTTVNASKELLLCYLI 420
WG-2.331_MUTANT_SEN2 --------------YSVIAELVDDNFEGSLRRPLSWMSLSGLNRTTVSASK-------- 255
WG-2.331_WT_SEN2     --------------YSVIAELVDDNFEGSLRRPLSWMSLSGLNRTTVNASK-------- 255
                                   **********************************.***

DT40_SEN2           KPSDMTEEEMATPECLKRIKVQELIVTRWVSSRERSEQDDL 461
WG-2.331_MUTANT_SEN2 ---------------------ELIVTRWVSSRERSEQDDL 275
WG-2.331_WT_SEN2     ---------------------ELIVTRWVSSRERSEQDDL 275
                                          *******************
```

Figure 2. Alignment of TSEN2 protein of normal and mutant Wg-2.331 samples in reference to DT40. Seven amino acid substitutions based on sequencing results were identified in *wg-2/wg-2* mutant samples. The normal and mutant TSEN2 proteins were then aligned to the NCBI reference protein for TSEN2 (DT40_SEN2) to determine possible similarities. The alignment found that 3 of the substituted amino acids were present in the TSEN2 reference [Y149D, A211T, and A214T, in (a) boxes]. For the remaining substitutions, 3 were unique to the mutant samples [I75V, A232G, and N459S, in (b) boxes], and one was unique in both normal and mutant samples in comparison to the TSEN2 reference [M356V, in (c) box]. One site was polymorphic for amino acid across all 3 samples [in (d) box]. The NCBI reference was sequenced from the DT40 cell line derived from Single Comb White Leghorn bursal lymphocytes (Caldwell et al., 2005, NP_001025765.1).

Candidate Gene TSEN2: Haplotype Analysis

To confirm the amino acid substitutions observed in *TSEN2*, 2 additional homozygous normal (+/+) and 2 homozygous mutant (*wg-2/wg-2*) individuals were sequenced for the entire *TSEN2* gene using overlapping primer pairs (Supplemental Table S4). Samples were selected based on their confirmed phenotypes and SNP390 genotypes.

Complete sequencing of the 10 kb *TSEN2* gene (and flanking regions) revealed 2 distinct genotype-specific haplotypes. The mutant (*wg-2/wg-2*) haplotype retained the polymorphisms observed in the original sequencing study along with newly identified variants. Surprisingly, one homozygous normal (SNP390 +/+) sample was heterozygous for the 2 *TSEN2* haplotypes. Thus, it seemed likely that recombination had occurred between *TSEN2* and marker SNP390. To verify this putative recombination event, 29 additional samples were sequenced using one primer pair (AW-45, see Supplemental Table S4).

Of the 29 samples (normal, carrier, and mutant birds) sequenced at this smaller section, 6 exhibited evidence of a recombination event (Supplemental Table S7). One of these, 2A1, was then sequenced across *TSEN2* in full to determine if this event indicated loss of linkage between the phenotype and the previously identified mutant-associated *TSEN2* polymorphisms. Sequence from this sample, a phenotypically confirmed normal embryo heterozygous for SNP390 (+/*wg-2*), exhibited homozygous polymorphisms in *TSEN2* matching those in the mutant individuals (Figure 2). The presence of these polymorphisms in a phenotypically normal embryo indicates that they are not linked to the mutant phenotype. Also based on these results, the *wg-2* causative region was reduced to a CR_{max} of 267 kb, from a *TSEN2* SNP in the terminal exon of the gene (referred to as AW-45, amplified by primer set AW-45) to SNP MK-21, or 4,927,330 to 5,194,358 as per the galGal5 assembly (Figure 1C), and therefore 2 of the candidate genes, *TSEN2* and *PPARG* (5′ of the recombination event), were eliminated from the

wg-2 region of interest leaving 4 known genes (*MKRN2, RAF1, RAB43,* and *CNBP*).

Linkage Analysis in New Generations of Carriers

Over several years subsequent to the *TSEN2* analysis, the population was maintained by the breeding of carriers determined by SNP390 genotyping. We therefore decided to examine the ends of the linked region in SNP390 +/*wg-2* individuals starting in 2014, to assess if recombination events had occurred in the intervening years. Several putative recombination events were identified in the 2014 and then 2015 carrier birds, and these were bred to perpetuate the Wg-2.331 line. An additional 19 SNP were genotyped to assess the recombination breakpoints (Supplemental Table S5). These SNP were derived from a capture array (Robb and Delany, 2012) and independently observed by other researchers as shown by their listing in the Beijing Genome Institute database. A new genome assembly (galGal5 in 2015) became available, such that the prior causative region was shown to span a 267 kb region and included a fourth validated gene, *RAB43*. The new recombination events observed in carriers of the 2014 and 2015 generations indicated a CR_{max} of 226,637 bp in the galGal5 assembly, from marker IY-13 to IY-3 (Figure 1D). This region excluded *MKRN2*, limiting the list of validated candidate genes to 3 (*RAF1, CNBP,* and *RAB43*, Figure 1D).

DISCUSSION

Prior to the use of the Illumina 3 K SNP chicken array (Muir et al., 2008a; Muir et al., 2008b) and application of other genomics tools described here, the *wg-2* mutation was genetically characterized only by its autosomal recessive mode of inheritance (Somes, 1990). By employing the chicken 3 K SNP array we established an initial region of interest of ca. 7 Mb mapping to GGA 12 and affirmed the autosomal mode of inheritance (Gitter, 2006). An important additional benefit of these results was the discovery of marker SNP390, which to date continues to exist in complete LD with the *wg-2* mutation. SNP390 was incorporated in the Wg-2.331 congenic inbred line breeding program as a predictor of *wg-2* genotype. This marker is a significant aid for the long-term perpetuation of the line, as the predictive genotyping conserves limited resources by reducing animal numbers, staff time, and the related costs. However, caution must be exercised in using a single or even several markers alone in predicting carriers; such predicted carriers must be test-mated routinely to ensure recombination does not disrupt the association, which, if it occurs without awareness, could result in loss of the mutation. The alternative dilemma is that by using only a single marker (e.g., SNP390) in selecting carriers for breeding during the early years of a project such as

this, the breeders chosen carry a region surrounding the polymorphism that is not disrupted by recombination. Thus, use of markers for breeding schemes must be part of a comprehensive and balanced approach depending on the scientific, genetic, and breeding goals.

In this study, several complementary methods were employed toward our long-term aim of identifying the causative element of the *wg-2* phenotype. SNP genotyping coupled with fine-mapping was used to take advantage of naturally occurring recombination events to reduce the region, while direct study of gene sequences prioritizes mutations that may be highly disruptive. Direct sequencing, SNP genotyping, predicting amino acid substitutions resulting from polymorphisms, and comparing polymorphisms among different organisms have all served to refine the potential causative region to a maximal size of 227 kb, containing only 3 confirmed genes as per the latest assembly (galGal5). An important consideration for studies such as this one and for candidate gene analysis in general is the quality of the genome assembly. The GGA 12 region of interest as per the galGal5 assembly has no evidence of sequencing gaps and exhibits conserved synteny for *PPARG, TSEN2, MKRN2,* and *RAF1* with the human (GRCh37/hg19) and mouse (GRCm38/mm10) assemblies (http://genome.ucsc.edu/) of HSA 3 and MMU 6, respectively. *CNBP* and *RAB43* are also syntenic in both human and mouse, but are not as proximal to the other genes as in the chicken assembly. While the first 4 genes map to the p arm of HSA 3 in human, *CNBP* and *RAB43* are on the q arm; in mouse *Cnbp* and *Rab43* are actually 5′ of the other genes in MMU 6. There are also 3 genes that the UCSC Genome Browser displays on the Ensembl Gene Prediction track that may become fully validated in a future chicken genome assembly, as occurred with *RAB43* during the course of this study. The predicted genes have sequence homology with known genes in other organisms, but have not been fully reviewed and validated in chicken. However, all 3 predicted genes have EST and are therefore likely to be real. Table 2 compares the known genomic attributes of the 3 validated genes and the 3 predicted genes in chicken, mouse, and human. Both validated and predicted genes of the *wg-2* causative region are discussed briefly below with a comparative overview as to developmental function(s).

RAF1 (v-*raf*-1 murine leukemia viral oncogene homolog) is a serine-threonine kinase that belongs to the *RAS-MAPK* signaling pathway (http://www.uniprot.org/uniprot/P04049). Mutations in *RAF1* are associated with 2 human developmental disorders, LEOPARD syndrome and Noonan syndrome, which are both characterized by cardiac and facial abnormalities with additional anomalies unique to each disorder (Sarkozy et al., 2008; Tidyman and Rauen, 2008). RAF1 regulates a member of the fibroblast growth factor (**FGF**) signaling family (MacNicol et al., 1993; Fongang and Kudlicki 2013). Notably, FGF8 is an inducer of chick limb development (Crossley et al., 1996).

Table 2. Comparative genomic attributes of genes in the 227 kb chicken causative region of *wingless-2*.

	Start Location (Mb)			Gene Size (bp)		
Validated genes	Chicken, GGA 12[1]	Human, HSA 3[2]	Mouse, MMU 6[3]	Chicken	Human	Mouse
RAF1	4.95[4]	12.58	115.62	27,932	80,626	58,063
CNBP	5.04	129.17	87.84	9269	14,439	8492
RAB43	5.09	129.09	87.79	34,582	34,582	22,927
Predicted Genes						
ISY1	5.05	129.13	87.82	9211	31,674	20,312
CECR5	5.07	17.14 (HSA 22)	120.51	7163	27,767	21,806
EFCC1/CCDC48	5.15	129	87.73	42,751	39,113	25,043

[1]Gallus˙gallus-5.0/galGal5.
[2]GRCh38/hg38.
[3]CRCm38/mm10.
[4]Only 13.5 kb of RAF1 are included in the causative region (see Figure 1D).

Cellular nucleic acid binding protein (**CNBP**) is hypothesized to be involved in tissue remodeling with a controlling role in both cell death and proliferation rates (Calcaterra et al., 2010). CNBP expression in a gastrula stage (about 24 h of incubation) chick embryo was reported to localize to the neuroectoderm of the prospective forebrain (Abe et al., 2006). By ca. 3 to 3.5 d of incubation, CNBP expression localized to the forebrain and midbrain with low expression in the tail, heart, and wing buds (Abe et al., 2006). CNBP knockdown caused truncation of the forebrain in chick (Abe et al., 2006) and mouse (Chen et al., 2003) embryos. Knockdown of this binding protein in zebrafish established its role in pharyngeal and craniofacial cartilaginous structures derived from the neural crest (**NC**). Specifically, morpholino-dependent depletion of the protein caused an increase in NC cell apoptosis and a decrease in proliferation (Weiner et al., 2011).

RAB43 is a member of the RAS oncogene family of small GTPases, which regulate intracellular membrane trafficking involving vesicles (formation, movement, tethering, and fusion). Human RAB43 is required for maintaining the structure/function of the Golgi complex (Haas et al., 2007; http://www.uniprot.org/uniprot/Q86YS6). *Rab43* knockout mice are relatively normal, and no developmental impacts were reported (Kretzer et al., 2016).

The 3 predicted genes within the 227 kb *wg-2* region of GGA 12 are *ISY1* [NM_020701.2], *CECR5* [NM_017829.5], and *CCDC48* [NM_024768.2]. In *Saccharomyces cerevisiae*, *ISY1* (Interactor of Syf1p) was identified as part of a spliceosome complex (http://www.uniprot.org/uniprot/Q9ULR0). A knockout (isy1Δ) resulted in lower pre-mRNA splicing efficiency indicating a function in pre-mRNA splicing (Dix et al., 1999). This predicted gene includes the SNP390 marker within a predicted intron. *CECR5* (Cat Eye Syndrome Chromosome Region, Candidate 5) is a candidate gene for the human Schmid-Fraccaro Syndrome (Cat Eye Syndrome) resulting from partial duplication or trisomy of HSA 22. The affected individuals express variable phenotypes involving skeletal, craniofacial, and organ systems (Footz et al., 2001; http://www.uniprot.org/uniprot/Q9BXW7). The predicted gene *CCDC48* (Coiled-coil domain containing

48), also known as *EFCC1* (EF-Hand and coiled-coil domain containing 1), is a protein-coding gene with Gene Ontology annotation related to calcium ion binding, with little known about the gene otherwise (http://www.uniprot.org/uniprot/Q9HA90).

In summary, the *wg-2* mutation was mapped to GGA 12 within a maximal 227 kb candidate region. This is a conservative estimate of the region, as the edges are defined by the outermost SNP that have been observed to be unlinked to the mutation. The distance between the unlinked SNP and the next instances of linkage is included in the CR_{max} to ensure that the causative mutation is included. The current size of the region is very small given the estimated recombination rate for chicken chromosomes (median value of 2.8 cM/Mb on macro- and 6.4 cM/Mb on microchromosomes, International Chicken Genome Sequencing Consortium, 2004). Although fine-mapping is unlikely to provide anything more than marginal reductions in the size of the causative region in the future, we continue to genotype at the margins to detect any recombination that might occur. Over the course of our work, which included fine-mapping and sequencing along with availability of improved assemblies, we eliminated 4 candidate genes from an original list of 7 genes. Thus, we have *RAF1, CNBP, RAB43,* and the 3 predicted genes for experimental consideration. As the predicted genes have reported EST, they are likely real and cannot be excluded. Our current research involves the development and analysis of a new 300 kb region capture array designed to obtain high-quality sequence reads for the GGA 12 *wg-2* region of interest and small flanking regions across all coding and non-coding DNA. Pooled DNA from mutants was sequenced using this capture array technology to provide an initial assessment of overall variation in the region (Robb and Delany, 2012). The new array employs 2 wild type, 4 carrier, and 18 mutant individuals to improve our ability to distinguish potentially causative variation from the high degree of polymorphisms observed. A causative variant must be exclusive to and shared by all *wg-2* mutant homozygotes, found heterozygous in carriers and absent in normal homozygotes. This new capture array effort will produce a comprehensive understanding of the *wg-2* region by elucidating all variation, including SNP, indels,

gaps, duplications, and any structural rearrangements, even in non-genic regions (e.g., promoters, enhancers, introns, and intergenic regions). Going forward, this dataset will enumerate the comprehensive list of those variants unique to, and potentially causative of, *wg-2*. Ultimately, to establish the cause of the phenotype, sequence-based approaches will be paired with study of gene expression differences between *wg-2* mutant and normal siblings.

SUPPLEMENTARY DATA

Supplementary data are available at *Poultry Science* online.

Supplemental Table S1. PCR primers utilized for fine mapping the 2 Mb *wingless-2* causative region to 842 kb of GGA 12.

Supplemental Table S2. Informative SNP derived from sequencing (Supplemental Table S1) showing linkage to Wg-2.331.

Supplemental Table S3. PCR primers for exons of candidate genes (*PPARG, TSEN2, MKRN2, RAF1, CNBP*) in 842 kb causative region (on GGA 12).

Supplemental Table S4. PCR primers for *TSEN-2* (on GGA 12) sequencing of UCD Wg-2.331 individuals.

Supplemental Table S5. PCR primers to fine map the 267 kb *wingless-2* causative region to 227 kb of GGA 12.

Supplemental Table S6. Informative SNP in showing linkage to Wg-2.331 from a 3 K array (Muir et al., 2008a).

Supplemental Table S7. Evidence for recombination between *TSEN2* and SNP390. *TSEN2* sequencing was conducted in 29 individuals (tables below), resulting in the identification of a recombination between one SNP locus (AW-45) of *TSEN2* and marker SNP390 in 6 normal phenotype individuals (table at right). The pedigree of those individuals shows they were from the same family, suggesting inheritance of a single recombined chromosome present in one of the parents (parents were not sequenced).

ACKNOWLEDGMENTS

REFERENCES

Abe, Y., W. Chen, W. Huang, M. Nishino, and Y.-P. Li. 2006. CNBP regulates forebrain formation at organogenesis stage in chick embryos. Dev. Biol. 295:116–127.

Abplanalp, H., K. Sato, D. Napolitano, and J. Reid. 1992. Reproductive performance of inbred congenic Leghorns carrying different haplotypes for the major histocompatibility complex. Poult. Sci. 71:9–17.

Baba, T. W., and E. H. Humphries. 1984. Differential response to avian leukosis virus infection exhibited by two chicken lines. Virology. 135:181–188.

Calcaterra, N. B., P. Armas, A. M. Weiner, and M. Borgognone. 2010. CNBP: A multifunctional nucleic acid chaperone involved in cell death and proliferation control. IUBMB Life. 62: 707–714.

Caldwell, R. B., A. M. Kierzek, H. Arakawa, Y. Bezzubov, J. Zaim, P. Fiedler, S. Kutter, A. Blagodatski, D. Kostovska, M. Koter, J. Plachy, P. Carninci, Y. Hayashizaki, and J. M. Buerstedde. 2005. Full-length cDNAs from chicken bursal lymphocytes to facilitate gene function analysis. Genome Biol. 6:R6.

Carrington, J. L., and J. F. Fallon. 1984. Evidence that the ectoderm is the affected germ layer in thewingless mutant chick embryo. J. Exp. Zool. 232:297–308.

Chen, W., Y. Liang, W. Deng, K. Shimizu, A. M. Ashique, E. Li, and Y. P. Li. 2003. The zinc-finger protein CNBP is required for forebrain formation in the mouse. Development. 130:1367–1379.

Crossley, P. H., G. Minowada, C. A. MacArthur, and G. R. Martin. 1996. Roles for FGF8 in the induction, initiation, and maintenance of chick limb development. Cell. 84:127–136.

Dix, I., C. Russell, S. B. Yehuda, M. Kupiec, and J. D. Beggs. 1999. The identification and characterization of a novel splicing protein, Isy1p, of Saccharomyces cerevisiae. RNA. 5:360–368.

Engels, W. R. 1993. Contributing software to the internet: The Amplify program. Trends Biochem. Sci. 18:448–450.

Footz, T. K., P. Brinkman-Mills, G. S. Banting, S. A. Maier, M. A. Riazi, L. Bridgland, S. Hu, B. Birren, S. Minoshima, N. Shimizu, H. Pan, T. Nguyen, F. Fang, Y. Fu, L. Ray, H. Wu, S. Shaull, S. Phan, Z. Yao, F. Chen, A. Huan, P. Hu, Q. Wang, P. Loh, S. Qi, B. A. Roe, and H. E. McDermid. 2001. Analysis of the cat eye syndrome critical region in humans and the region of conserved synteny in mice: A search for candidate genes at or near the human chromosome 22 pericentromere. Genome Res. 11:1053–1070.

Fongang, B., and A. Kudlicki. 2013. The precise timeline of transcriptional regulation reveals causation in mouse somitogenesis network. BMC Dev. Biol. 13:42.

Gitter, C. L. 2006. Genetic mapping of seven developmental mutations found in the domestic chicken (Gallus gallus) by analysis of single nucleotide polymorphisms. M.S. Thesis. Univ. California, Davis.

Haas, A. K., S. Yoshimura, D. J. Stephens, C. Preisinger, E. Fuchs, and F. A. Barr. 2007. Analysis of GTPase-activating proteins: Rab1 and Rab43 are key Rabs required to maintain a functional Golgi complex in human cells. J. Cell Sci. 120(Pt 17):2997–3010.

Hamburger, V., and H. L. Hamilton. 1992. A series of normal stages in the development of the chick embryo. Dev. Dyn. 195:231–272.

International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 432:695–716.

Kretzer, N. M., D. J. Theisen, R. Tussiwand, C. G. Briseño, G. E. Grajales-Reyes, X. Wu, V. Durai, J. Albring, P. Bagadia, T. L. Murphy, and K. M. Murphy. 2016. RAB43 facilitates cross-presentation of cell-associated antigens by CD8α+ dendritic cells. J. Exp. Med. 213:2871–2883.

MacNicol, A. M., A. J. Muslin, and L. T. Williams. 1993. Raf-1 kinase is essential for early Xenopus development and mediates the induction of mesoderm by FGF. Cell. 73:571–583.

Muir, W. M., G. K. Wong, Y. Zhang, J. Wang, M. A. M. Groenen, R. P. M. A. Crooijmans, H.-J. Megens, H. M. Zhang, J. C. McKay, S. McLeod, R. Okimoto, J. E. Fulton, P. Settar, N. P. O'Sullivan, A. Vereijken, A. Jungerius-Rattink, G. A. A. Albers, C. Taylor Lawley, M. E. Delany, and H. H. Cheng. 2008. Review of the

initial validation and characterization of a 3 K chicken SNP array. Worlds Poult. Sci. J. 64:219–226.

Muir, W. M., G. K.-S. Wong, Y. Zhang, J. Wang, M. A. M. Groenen, R. P. M. A. Crooijmans, H.-J. Megens, H. Zhang, R. Okimoto, A. Vereijken, A. Jungerius, G. A. A. Albers, C. Taylor Lawley, M. E. Delany, S. MacEachern, and H. H. Cheng. 2008. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. Proc. Natl. Acad. Sci. 105:17312–17317.

Pisenti, J. M. 1995. Genetic and environmental influences on the expression of the mutation wingless-2 in the chick embryo. PhD Diss. Univ. California, Davis.

Pisenti, J. M., M. E. Delany, R. L. Taylor, U. K. Abbott, H. Abplanalp, J. A. Arthur, M. R. Bakst, C. Baxter-Jones, J. J. Bitgood, F. A. Bradley, K. M. Cheng, R. R. Dietert, J. B. Dodgson, A. M. Donoghue, A. B. Emsley, R. J. Etches, R. R. Frahm, R. J. Gerrits, P. F. Goetinck, A. A. Grunder, D. E. Harry, S. J. Lamont, G. R. Martin, P. E. McGuire, G. P. Moberg, L. J. Pierro, C. O. Qualset, M. A. Qureshi, F. T. Shultz, and B. W. Wilson. 1999. Avian genetic resources at risk: An assessment and proposal for conservation of genetic stocks in the USA and Canada. GRCP Report No. 20. Univ. California Division of Agriculture and Natural Resources, Genetic Resources Conservation Program, Davis CA. http://grcp.ucdavis.edu/publications/index.htm.

Robb, E. A., C. L. Gitter, H. H. Cheng, and M. E. Delany. 2011. Chromosomal mapping and candidate gene discovery of chicken developmental mutants and genome-wide variation analysis of MHC congenics. J. Hered. 102:141–156

Robb, E. A., and M. E. Delany. 2012. Case study of sequence capture enrichment technology: Identification of variation underpinning developmental syndromes in an amniote model. Genes. 3:233–247.

Sarkozy, A., M. C. Digilio, and B. Dallapiccola. 2008. Leopard syndrome. Orphanet. J. Rare. Dis. 3:13.

Somes, R. G. 1990. Lethal mutant traits in chickens. Poultry Breeding and Genetics. Pages 293–331. R. D. Crawford, ed. Elsevier Publishers, BV, Amsterdam.

Tidyman, W. E., and K. A. Rauen. 2008. Noonan, Costello and cardio-facio-cutaneous syndromes: Dysregulation of the Ras-MAPK pathway. Expert Rev. Mol. Med. 10:e37.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen. 2012. Primer3–New capabilities and interfaces. Nucleic. Acids. Res. 40:e115–e115.

Waters, N. F., and J. H. Bywaters. 1943. A lethal embryonic wing mutation in the domestic fowl. J. Hered. 34:213–217.

Weiner, A. M. J., M. A. Sdrigotti, R. N. Kelsh, and N. B. Calcaterra. 2011. Deciphering the cellular and molecular roles of cellular nucleic acid binding protein during cranial neural crest development. Dev. Growth Differ. 53:934–947.

Zwilling, E. 1956. Interaction between limb bud ectoderm and mesoderm in the chick embryo. II. Experimental limb duplication. J. Exp. Zool. 132:173–187.